

```
QY      99 aaaccacagtagaaaa 117
        ||| ||| ||| ||| ||| |||
Db      244 AAACACAGTAGAAAAA 226
```

**TITLE** Use of BAC End Sequences from Caltech Libraries for sequence-ready map building





Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 ttttaataaaattttgaa 306  
 Db 28 tttcataataaattgttga 48

RESULT 10  
 SOYMOD26A 5831 bp DNA PLN 14-OCT-1993  
 LOCUS Glycine max cv. Dare nodulin 26 gene fragment.  
 DEFINITION M94764  
 ACCESSION M94764.1 GI:170029  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Glycine max (cultivar Dare) DNA.

REFERENCE 1 (bases 1 to 5831)  
 AUTHORS Miao, G.H. and Verma, D.P.  
 TITLE Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants  
 JOURNAL Plant Cell 5 (7), 781-794 (1993)  
 MEDLINE 93372569  
 FEATURES Location/Qualifiers  
 1..5831  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"

BASE COUNT 2029 a 824 c 877 g 2101 t  
 ORIGIN

Query Match 6.8%; Score 21; DB 7; Length 5831;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 acctgcaagctttgtgttg 101  
 Db 2648 ACCTGCAAGCTTTGTGTG 2668

RESULT 11  
 AC025334/C 71804 bp DNA HTG 08-MAR-2000  
 LOCUS Homo sapiens chromosome 2 clone RP11-358M3 map 2, LOW-PASS SEQUENCE  
 DEFINITION  
 SAMPLING.  
 AC025334  
 AC025334.1 GI:7210008  
 VERSION  
 KEYWORDS HTG; HTGS-PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 71801)  
 AUTHORS Birren, B., Linton, L., Nussbaum, C. and Lander, E.  
 TITLE Homo sapiens chromosome 2, clone RP11-358M3  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 71801)  
 AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, E., Boguski, M., Bork, P., Brown, A., Burkett, D., Campione, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collins, S., Collins, S., Cooke, P., Dear, P., Devereux, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hages, B., Heath, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroix, K., Lamazares, R., Landers, T., Lenoci, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,

TITLE  
 JOURNAL  
 COMMENT

Meldrum, J., Meneus, L., Mithova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, G.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 --- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 --- Project Information  
 Center project name: L6454  
 Center clone name: 358\_M3

NOTE: This record contains 84 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 746: contig of 746 bp in length  
 747 846: gap of 100 bp  
 847 1601: contig of 755 bp in length  
 1602 1701: gap of 100 bp  
 1702 2462: contig of 761 bp in length  
 2463 2562: gap of 100 bp  
 2563 3321: contig of 759 bp in length  
 3322 3421: gap of 100 bp  
 3422 4158: contig of 737 bp in length  
 4159 4258: gap of 100 bp  
 4259 5011: contig of 753 bp in length  
 5012 5111: gap of 100 bp  
 5112 5874: contig of 763 bp in length  
 5875 5974: gap of 100 bp  
 5975 6704: contig of 730 bp in length  
 6705 6804: gap of 100 bp  
 6805 7562: contig of 758 bp in length  
 7563 7662: gap of 100 bp  
 7663 8434: contig of 772 bp in length  
 8435 8534: gap of 100 bp  
 8535 9298: contig of 764 bp in length  
 9299 9398: gap of 100 bp  
 9399 10157: contig of 759 bp in length  
 10158 10257: gap of 100 bp  
 10258 10999: contig of 742 bp in length  
 11000 11099: gap of 100 bp  
 11100 11854: contig of 755 bp in length  
 11855 11954: gap of 100 bp  
 11955 12717: contig of 763 bp in length  
 12718 12817: gap of 100 bp  
 12818 13568: contig of 751 bp in length  
 13569 13668: gap of 100 bp  
 13669 14423: contig of 755 bp in length  
 14424 14523: gap of 100 bp  
 14524 15288: contig of 765 bp in length  
 15289 15388: gap of 100 bp  
 15389 16435: contig of 747 bp in length  
 16436 16235: gap of 100 bp

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	16.2	162	27	AA493099
2	55	12.4	300	90	W07101
3	55	12.4	451	39	AA417464
4	55	12.4	496	32	AA857673
5	55	12.4	497	71	AA328446
6	55	12.4	498	37	AI277373
7	55	12.4	504	39	AI439816
8	55	12.4	513	37	AI278440
9	55	12.4	525	37	AI435639
10	55	12.4	534	29	AA490602
11	55	12.4	551	27	AA480162
12	55	12.4	554	29	AA663472
13	55	12.4	554	29	AA439530
14	51	11.5	308	102	AA439530
15	51	11.5	420	47	AI983716
16	51	11.5	520	38	AI336216
17	51	11.5	572	79	AA674023
18	50	11.2	457	63	AA103030
19	50	11.2	556	73	AA499920
20	48	10.6	354	107	AA323354
21	47	10.6	454	98	AA018060
22	47	10.6	523	96	AA027318
23	47	10.6	526	111	AA069274
24	46	10.3	544	43	AI744949
25	45	10.1	485	95	AA0167057
26	45	10.1	489	110	AA0799581
27	45	10.1	622	120	BA48282
28	45	10.1	864	101	AA0342901
29	44	9.9	186	28	AA576792
30	44	9.9	344	70	AA302950
31	44	9.9	375	35	AI144036
32	44	9.9	420	21	AA085410
33	43	9.7	353	27	AA507837
34	43	9.7	667	22	AA196790
35	43	9.7	980	110	AA077348
36	42	9.4	385	27	AA490097
37	42	9.4	627	94	AA0114141
38	42	9.4	629	107	AA0635928
39	41	9.2	360	45	AA1858468
40	41	9.2	425	21	AA847029
41	41	9.2	454	105	AA0475892
42	40	9.0	299	96	AA0316521
43	40	9.0	374	94	AA0103431
44	40	9.0	403	31	AA831408
45	40	9.0	417	27	AA487412

## ALIGNMENTS

RESULT 1  
LOCUS AA493099/c  
DEFINITION 162 bp mRNA  
EST 19-AUG-1997  
n996d01.s1 NCI-CGAP\_Thy1 Homo sapiens cDNA clone IMAGE:942625  
Similar to contains Alu repetitive element; contains element M99

NCBI repetitive element ;, mRNA sequence.  
AA493099  
VERSION AA493099.1 GI:2222940  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1798572.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1350  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: [www-bio.illn.gov/bdrrp/image/image.html](http://www-bio.illn.gov/bdrrp/image/image.html)  
Insert length: 451 Std Error: 0.00  
Seq primer: -40m3 fwd. RT from Amersham  
High quality sequence stop: 153.  
Location/Qualifiers  
1. 162  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:942625"  
/clone\_lib="NCI-CGAP\_Thy1"  
/tissue\_type="thyroid"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; mRNA made from invasive thyroid tumor, CDNA made by oligo-dr priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

## FEATURES

BASE COUNT 47 a 36 c 50 g 29 t  
ORIGIN  
Query Match 16.2%; Score 72; DB 27; Length 162;  
Best Local Similarity 100.0%; Pred. No. 6.3e-24;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 tttagacagcgggttcgcacacttcgcagcgtgttcgaactcctgacctgtg 334  
DB 162 tttagacagcgggttcgcacacttcgcacacttcgcacacttcgttgaactcctgacctgtg 103  
QY 335 attaccacacct 346  
DB 102 attaccacacct 91

RESULT 2  
LOCUS W07101/c  
DEFINITION 300 bp mRNA  
EST 25-APR-1996  
IMAGE:300107.5' similar to contains Alu repetitive element; contains element M922 repetitive element ;, mRNA sequence.  
ACCESSION W07101  
VERSION W07101.1 GI:1281414  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 300)  
Hillier, L., Clark, N., Dubuque, J., Elliston, K., Hawkins, M.,

